Robust Estimation of Large Panels with Factor Structures

Marco Avarucci & Paolo Zaffaroni

To cite this article: Marco Avarucci & Paolo Zaffaroni (2022): Robust Estimation of Large Panels with Factor Structures, Journal of the American Statistical Association, DOI: 10.1080/01621459.2022.2050244

To link to this article: https://doi.org/10.1080/01621459.2022.2050244

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

View supplementary material

Published online: 11 Apr 2022.

Submit your article to this journal

Article views: 144

View related articles

View Crossmark data
Robust Estimation of Large Panels with Factor Structures

Marco Avarucci\textsuperscript{a} and Paolo Zaffaroni\textsuperscript{b,c}

\textsuperscript{a}University of Glasgow, Glasgow, UK; \textsuperscript{b}Imperial College Business School, London, UK; \textsuperscript{c}University of Rome La Sapienza, Rome, Italy

\textbf{ABSTRACT}

This article studies estimation of linear panel regression models with heterogeneous coefficients using a class of weighted least squares estimators, when both the regressors and the error possibly contain a common latent factor structure. Our theory is robust to the specification of such a factor structure because it does not require any information on the number of factors or estimation of the factor structure itself. Moreover, our theory is efficient, in certain circumstances, because it nests the GLS principle. We first show how our unfeasible weighted-estimator provides a bias-adjusted estimator with the conventional limiting distribution, for situations in which the OLS is affected by a first-order bias. The technical challenge resolved in the article consists of showing how these properties are preserved for the feasible weighted estimator in a double-asymptotics setting. Our theory is illustrated by extensive Monte Carlo experiments and an empirical application that investigates the link between capital accumulation and economic growth in an international setting. Supplementary materials for this article are available online.

1. Introduction

This article studies estimation of linear panel regression models with heterogeneous coefficients using a class of weighted least squares estimators, when both the regressors and the error possibly contain a common latent factor structure. Factor models represent one of the most popular and successful ways to capture cross-sectional and temporal dependence, especially when facing a large number of units (N) and time periods (T). However, in our context, the possibility of a common factor structure in both regressors and error leads to an endogeneity problem, making estimation of panel data models by ordinary least squares (OLS) invalid.

In principle, the endogeneity would be trivially resolved if one could hypothetically observe the latent common factors and use them in the OLS regression. Second, an efficiency improvement over the OLS estimator can be achieved if the true variance-covariance matrix of the idiosyncratic regression error is known, leading to the generalized least squares (GLS) estimator. Both of these steps are not feasible in practice, as we do not observe the latent common factors nor do we have the covariance matrix of the idiosyncratic error.

The contribution of this article is to show how to achieve these two goals, despite the latency of the factors and of the variance-covariance matrix of the idiosyncratic regression error. Unlike other approaches, we do not advocate preliminary estimation of the factor structure and, instead, we treat factors and their loadings as nuisance parameters rather than objects of interest. In particular, we construct a feasible weighted estimator that resolves the endogeneity without any knowledge of the common factors, qualifying our methodology as robust. This appears to be the strongest feature of our methodology.

Moreover, the statistical properties of our estimator are preserved under several circumstances of interest that can arise in the data.\textsuperscript{1} This is relevant as many existing estimation procedures designed to tackle the endogeneity problem in panel regressions, such as the popular common correlated effects (hereafter CCE) estimator of Pesaran (2006), become invalid, that is, first-order biased, when any of the above circumstances apply. Moreover, the large majority of the contributions do not allow heterogeneity of the regression coefficients (see, e.g., Bai 2009). Resolving these problems, that is constructing an estimator robust to endogeneity when alternative procedures fail and allowing for heterogeneous slope coefficients, represents our advancement in the literature of panel estimation with interactive fixed effects.

Our work also advances the literature of GLS estimation along two dimensions. In fact, first, under specific circumstances on the idiosyncratic error, the weighting scheme adopted by our estimator achieves full efficiency, meaning that it is asymptotically equivalent to a feasible GLS estimator. We qualify this property as quasi-efficiency, given that these circumstances on the idiosyncratic error are not generally warranted. Second, our work also contributes to the literature that demonstrates how efficient estimation techniques not only lead to an improvement of precision but, most importantly, may resurrect the required asymptotic properties, in terms of bias, rate of convergence, and distribution, when these fail in inefficient approaches (see Phillips 1991; Robinson and Hidalgo 1997).
Section C of the supplementary materials offers a complete overview of these two vast streams of literature on estimation of panel with interactive fixed effects and efficient estimation of regression models.

We first consider the unfeasible weighted least squares estimator for the regression coefficients, assuming that both the covariance matrix of the idiosyncratic error and the latent factors are known. This unfeasible estimator is asymptotically equivalent to the (unfeasible) GLS estimator applied to the covariance matrix of the idiosyncratic error and the latent factor for the regression coefficients, assuming that both the regression models.

An overview of these two vast streams of literature on estimation bias weight matrix maintains the asymptotic orthogonality to sense of element-by-element convergence). Does not converge to the unfeasible weighting matrix (in the scheme is invalid, and, indeed, the feasible weighting matrix is that the feasible weighting matrix can be constructed using residuals. Lack of consistency of the OLS estimator for the coefficients is 

2 Our simulation exercise shows that these two unfeasible estimators are asymptotically equivalent when T is large.

2. Methodology: Model and Estimators

The notation adopted throughout the article is collected in Appendix B, supplementary materials but we also define quantities throughout the article when needed. Assume that the observed variables obey a linear regression model with S common regressors \( d_i \) and K heterogeneous regressors \( x_{it} = [x_{it1} \ldots x_{itK}]' \). Following the convenient specification put forward by Pesaran (2006), the model for the ith unit can be expressed, in matrix form, as

\[
y_i = D\alpha_i + X_i\beta_i + u_i,
\]

(2.1)

for an observed \( T \times 1 \) vector \( y_i = [y_{it1} \ldots y_{iT}]' \), an observed \( T \times S \) matrix \( D = [d_{1i} \ldots d_{Ti}]' \) of common regressors, an observed \( T \times K \) matrix \( X_i = [x_{i1} \ldots x_{iT}]' \) of heterogenous regressors, an unobserved \( T \times 1 \) vector \( u_i = [u_{it1} \ldots u_{iT}]' \) of regression errors, and slope coefficients \( \alpha_i \) and \( \beta_i \). The presence of common observed regressors are not necessary for our results, and, in fact, complicate somewhat the asymptotic analysis but could be relevant in practical applications, so we provide a unified theory. In turn, the error vector satisfies the following latent factor structure:

\[
u_i = F\varepsilon_i + \varepsilon_i, \quad \text{with } E\varepsilon_i = 0 \quad \text{and } \quad \Xi_i := E\varepsilon_i\varepsilon_i' ,
\]

(2.2)

for an unobserved \( M \times 1 \) vector of factor loadings \( b_i \), an unobserved \( T \times M \) matrix of common factors \( F = \{f_1, \ldots, f_M\}' \), and an unobserved \( T \times 1 \) vector of idiosyncratic errors \( \varepsilon_i = [\varepsilon_{i1} \ldots \varepsilon_{iT}]' \). We postulate that the heterogeneous regressors satisfy

\[
X_i = D\Delta_i + F\Gamma_i + V_i ,
\]

(2.3)

for an unobserved \( S \times K \) matrix of factor loadings \( \Delta_i = [\delta_{i1} \ldots \delta_{iS}] \) with \( \delta_{it} = [\delta_{it1} \ldots \delta_{itK}]' \), an unobserved \( M \times K \) matrix of factor loadings \( \Gamma_i = \{\gamma_{i1} \ldots \gamma_{im}\}' \) with \( \gamma_{it} = [\gamma_{it1} \ldots \gamma_{itK}]' \), and an unobserved \( T \times K \) matrix of idiosyncratic errors \( V_i = [v_{i1} \ldots v_{iT}]' \) with \( v_{it} = [v_{it1} \ldots v_{itK}]' \). Formulation (2.3) can be substantially relaxed, for instance without imposing a linear factor structure in the \( X_i \), although certainly at the cost of less primitive assumptions. However, (2.3) represents a simple yet powerful way to allow dependence between the regressors and the errors of (2.1) through the latent factor structure.\(^{3}

\(^{3}\)As explained below, by exploiting the linearity assumption embedded in (2.2)–(2.3), the weighting scheme of our FWLS estimator implies that the idiosyncratic components of \( X_i \) and \( u_i \) that is \( V_i \) and \( \varepsilon_i \), respectively, dominate the limiting properties of the FWLS estimator, as the terms involving \( F \) become asymptotically negligible.
this respect, an important generalisation consists of allowing the $X_i$ to depend \textit{nonlinearly} on the $F$, a simple example of which being when the $X_i$ are function of polynomials in the $F$ (see Appendix F of the supplementary materials). The maintained condition here is that $K, S, M$ do not \textit{vary} with $T$ and $N$. However, most importantly, one does not need to know \textit{virtually} anything about $M$, not even an upper bound, as our methodology will work regardless of whether $M$ is either \textit{smaller, equal, or bigger} than $K$. \footnote{Our results can be extended to the case of $M$ (slowly) increasing with $T$ as a suitable rate.} Although model (2.1) is written as a single regression across time for a given unit $i$, we assume that a panel of observations $\{y_1, \ldots, y_N, X_1, \ldots, X_N\}$ is available and fully used within our methodology.

Our main objective is to estimate the heterogeneous slope coefficients $\beta_i$ of (2.1). We defer discussion on estimation of the slope coefficients $\beta_i$ of the common regressors to Section E of the supplementary materials. As explained below, throughout our analysis we always net out the effect of the common regressors by premultiplying the data by $M_D := I_T - D(D' D)^-1 D'$, where $I_T$ is the identity matrix of dimension $T$ and $A^+$ denotes the Moore–Penrose inverse of any matrix $A$. This allows us to avoid making any assumptions on $D_i$. Hence, one obtains

$$M_D y_i = \beta_i + M_D u_i.$$  

(2.4)

We consider three different estimators for the parameters $\beta_i$, namely the OLS and the \textit{unfeasible} and \textit{feasible} weighted least squares estimators, hereafter indicated as the WLS and FWLS estimators, respectively. Here we explain how to construct the estimators and, subsequently, provide the intuition on how the weighted estimators mitigate the endogeneity bias and, possibly, achieve efficiency.

Regarding the OLS estimator for $\beta_i$,$$
\beta_i^{OLS} := (X_i' M_D X_i)^{-1} X_i' M_D y_i. $$

(2.5)

We now describe the weighted least squares estimators.

Proceeding along the lines of (Magnus and Neudecker 1998, sec. 11—chap. 13) we define the WLS estimator, whose weighting matrix is based on taking the generalized inverse of $M_D S_i M_D$ with $S_i := FB' + \Sigma_i$, as follows$^5$

$$\beta_i^{WLS} := (X_i' M_D S_i M_D)^{-1} X_i' M_D y_i.$$  

(2.6)

The matrix $(M_D S_i M_D)$ is singular, leading to the composite expression for $\beta_i^{WLS}$ of (2.6), but by, Klemm K.2 in the supplemental materials, $(M_D S_i M_D)^{-1} = D_D (D_D' S_D D_D)^{-1} D_D'$, where $D_D$ spans the null space of $D$, namely the $T \times (T-S)$ full rank matrix such that $M_D = D_D D_D'$, where $D_D D_D' = I_{T-S}$. \footnote{To motivate the expression for $S_i$, notice that it can be obtained as the limit of the cross-sectional average of the $\sum_{i=1}^N b_i b_i'$, as $N \rightarrow \infty$, for the special homogeneity case $\Sigma_i = \Sigma$. Here $B$ defines the limit of $N^{-1} \sum_{i=1}^N b_i b_i'$, which we assume to be positive definite. Finally, pre and postmultiplication by $M_D$ is a consequence of netting out the effect of the common regressors $D$.}

By substitution, setting for simplicity,$$
y_i := D_i' y_i, \quad X_i := D_i' X_i, \quad \epsilon_i := D_i' \epsilon_i, \quad \mathcal{F} := D_i' F, \quad u_i := D_i' u_i, \quad V_i := D_i' V_i, $$

(2.7)

one obtains

$$\hat{\beta}_i^{WLS} = \left( X_i' D_i (D_i' S_D D_i)^{-1} D_i' X_i \right)^{-1} \left( X_i' D_i (D_i' S_D D_i)^{-1} D_i' y_i \right) = \left( \chi_i' s_i^{-1} \chi_i \right)^{-1} \chi_i' s_i^{-1} y_i, $$

(2.8)

where we set

$$s_i := D_i' S_i D_i = FB' + \chi_i \quad \text{with} \quad \chi_i := D_i' \Sigma_i D_i.$$  

(2.9)

This means that the WLS has now the more conventional expression (2.8) of a weighted estimator for the model

$$y_i = X_i \beta_i + u_i, \quad \text{with} \quad u_i = \mathcal{F} b_i + \epsilon_i, $$

(2.10)

without involving Moore–Penrose matrices, that is, with respect to a nonsingular weighting matrix $S_i^{-1}$. Premultiplying the data by $D_D$ reduces the sample size by $S$ time series observations since now the $y_i$ and the $X_i$ have $T - S$ rows. Likewise, considering again model (2.10), an equivalent representation to (2.5) is

$$\hat{\beta}_i^{OLS} = (X_i' X_i)^{-1} X_i' y_i.$$  

(2.11)

For the special case $S = 0$, that is when the common regressors $D$ are absent from the model, the above formulas continue to be valid, by setting $M_D = D_D = I_T$, yielding, for example, $\hat{\beta}_i^{WLS} = (X_i' X_i)^{-1} X_i' y_i$, and $\hat{\beta}_i^{OLS} = (X_i' S_i^{-1} X_i)^{-1} X_i' S_i^{-1} y_i$.

Considering model (2.10), our proposed FWLS estimator is given by

$$\hat{\beta}_i^{FWLS} := \left( \chi_i' \hat{S}_N^{-1} \chi_i \right)^{-1} \chi_i' \hat{S}_N^{-1} y_i,$$

(2.12)

where

$$\hat{S}_N := N^{-1} \sum_{i=1}^N \hat{u}_i \hat{u}_i', \quad \text{with} \quad \hat{u}_i := y_i - X_i \hat{\beta}_i^{OLS} = M_D u_i.$$  

(2.12)

Again, the special case for the FWLS estimator when $S = 0$ in (2.1), that is when no common regressors are present, is obtained replacing $D_D$ with $I_T$ in (2.11) and (2.12). Finally, our methodology also applies to the case of cross-sectional regressions with factor structure (see, Andrews 2005).\footnote{Given the analogies between GLS and Seemingly Unrelated Regression (SUR) estimator, one can in principle construct a SUR-type WLS estimator (see Appendix H, supplementary materials for details).}

3. Assumptions

We now present our assumptions, which, thanks to the specific model of (2.1)–(2.3), appear relatively primitive. Recall that, throughout our exposition, we always assume that $N > T$ holds, implied by condition $T/N = o(1)$, as demanded by our theory.

\footnote{Details are available upon request.}
Assumption 3.1 (idiosyncratic innovation \( \varepsilon_t \)). The \( N \times 1 \) vector \( \varepsilon_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tN})' \) satisfies the following equation:

\[
\varepsilon_t = R a_t, \quad \text{for } t = 1, \ldots, T,
\]

where the \( N \times N \) matrix of constants \( R = [r_{ij}] \), with \( r_{ij} \) not varying with \( N \) for given \( i \) and \( j \), has column- and row-norms satisfying \( \|R\|_{\text{row}} + \|R\|_{\text{col}} < \infty \), \( \min_i \sum_{j=1}^N |r_{ij}| > \kappa \) for some \( 0 < \kappa < \infty \) (not always the same), and the elements of the \( N \times 1 \) vector \( a_t = (a_{t1}, a_{t2}, \ldots, a_{tN})' \) follow a linear process,

\[
a_{it} = \sum_{s=0}^\infty \phi_{it} \eta_{is}, \quad \max_i \sum_{s=0}^\infty s^2 |\phi_{is}| < \infty, \quad \text{with } \phi_{01} = 1,
\]

where the elements of \( \eta_t = (\eta_{t1}, \ldots, \eta_{tN})' \) make an iid sequence across \( i \) and \( t \) with \( \mathbb{E} \eta_{it} = 0, \mathbb{E} |\eta_{it}|^2 < \infty \) and the smallest eigenvalue of their covariance matrix satisfying \( \lambda_1(\mathbb{E} \eta \eta') \geq \kappa > 0 \). Moreover, for every complex number \( z \) satisfying \( |z| \leq 1 \),

\[
\min_i |\phi_i(z)| > \kappa, \quad \text{where } \phi_i(z) := \sum_{s=0}^\infty \phi_{is} z^s.
\]

Remark 3.1. Assumption 3.1 is similar to Assumptions 1 and 2 in Pesaran and Tosetti (2011), and, with some variations, this form of cross-sectional and time dependence has been adopted also by Moon and Weidner (2015, 2017), and Onatski (2015). The assumption above turns out to be extremely convenient for establishing the asymptotic distribution of the WLS and FWLS estimators along the lines of Theorem 1 in Robinson and Hidalgo (1997). Notice that it implies that the sequence of the \( R \) matrices, as \( N \) increases, makes a nested family of matrices.

Remark 3.2. Assumption 3.1 implies that, for every \( 2 \leq h, \ell \leq 12 \),

\[
\max_{i_1} \max_{\ell_1, \ldots, \ell_{\ell-1}} N \sum_{i_1 \ldots i_{\ell-1} = 1} T \max_{t_1, \ldots, t_{\ell-1}} \sum_{t_1 \ldots t_{\ell-1} = 1} |\text{cum}_h(\varepsilon_{t_1 i_1}, \varepsilon_{t_2 i_2}, \ldots, \varepsilon_{t_{\ell-1} i_{\ell-1}})| < \infty,
\]

where the \( \text{cum}_h(\varepsilon_{t_1 i_1}, \varepsilon_{t_2 i_2}, \ldots, \varepsilon_{t_{\ell-1} i_{\ell-1}}) \) are the cumulants of order \( h \) of \( \varepsilon_{t_1 i_1}, \varepsilon_{t_2 i_2}, \ldots, \varepsilon_{t_{\ell-1} i_{\ell-1}} \).

Remark 3.3. By (Brockwell and Davis 1991, Lemma 4.5.3) (3.3) implies that the eigenvalues of the covariance matrices of \( a_t = (a_{t1}, \ldots, a_{tT})' \) are bounded and greater than \( \kappa \) for every \( j \). Easy calculations give \( \mathbb{E} a_t = \sum_{j=1}^N r_{ij}^2 \mathbb{E} a_j' \), implying that its smallest and largest eigenvalues satisfy \( \min_j \lambda_1(\mathbb{E} a_j) > \kappa \) and \( \max_j \lambda_T(\mathbb{E} a_j) < \infty \).

Assumption 3.2 (regressor innovation \( V_t \)). The sequence \( \{v_{itk}\} \) is covariance stationary, with zero mean, \( \max_{1 \leq k \leq h} \mathbb{E} |v_{itk}|^{14} < \infty \), and they satisfy, for every \( 2 \leq h, \ell, s \leq 14 \), and \( 2 \leq j \leq h \),

\[
\max_{k_1 \ldots k_s} \max_{t_1 \ldots t_\ell} N \sum_{i_1 \ldots i_s = 1} T \sum_{t_1 \ldots t_\ell = 1} (1 + t_1^2) |\text{cum}_h(v_{t_1 i_1 k_1}, \ldots, v_{t_\ell i_\ell k_\ell})| < \infty,
\]

where the \( \text{cum}_h(v_{t_1 i_1 k_1}, \ldots, v_{t_\ell i_\ell k_\ell}) \) are the cumulants of order \( h \) of \( v_{t_1 i_1 k_1}, \ldots, v_{t_\ell i_\ell k_\ell} \). Moreover, \( \min \lambda_1(\mathbb{E} v_{itk}') > \kappa \), where \( v_{itk}' = (v_{it1}, \ldots, v_{itk}) \).

Remark 3.4. Assumption 3.2 implies that \( T^{-1} V_t V_t' \xrightarrow{p} \Sigma V_t V_t' \), where \( \xrightarrow{p} \) means convergence in probability, with \( \min \lambda_1(\Sigma V_t V_t') > \kappa \). It follows that \( \max_i \| (V_t V_t'/T)^{-1} \| = O_p(1) \) as \( T \) diverges, where \( \| \cdot \| \) is the Frobenius norm.

Remark 3.5. The \( v_{itk} \) can be interpreted as the high-rank components of the regressors \( x_{itk} \) adopting Moon and Weidner (2015) terminology, as opposed to the \( D_t \), which represents the low-rank components. For instance, if for each \( k \), the \( v_{itk} \) are generated as \( \varepsilon_{itk} \) in Assumption 3.1, one obtains \( V_t = [v_{itk}]_{t,i=1}^{T,N} = O_p(\sqrt{\max(N,T)}) \) for every \( k \) (see the discussion in Moon and Weidner (2015, Appendix 1) and Onatski (2015)). In contrast, \( \sum_{t=1}^{T,N} \|D_t\|^2 \leq (N \|D\|^2)^{1/2} = O_p(\sqrt{NT}) \).

Assumption 3.3 (latent and observed factors). Set \( Z_t := [D_t', F_t'] = [z_{1t}, \ldots, z_{M+1,t}]' \) for \( 1 \leq t \leq T \) and \( 1 \leq r \leq M + S < \infty \). Then,

\[
\frac{Z'Z}{T} \xrightarrow{p} \Sigma_{Z'Z}, \quad \text{with } \Sigma_{Z'Z} := \begin{bmatrix} \Sigma_{DD} & \Sigma_{DF} \\ \Sigma_{FD} & \Sigma_{FF} \end{bmatrix} > 0,
\]

and \( \Sigma_{DD} > 0, \Sigma_{FF} > 0 \), where \( > \) means positive definitiveness. Moreover, we assume \( \mathbb{E} \|z_t\|^4 < \infty \), where \( z_t = [z_{1t}, \ldots, z_{M+1,t}]' \).

Remark 3.6. Equation (3.4) implies that \( \| (T^{-1/2} F_t' M_t F_t)^{-1} \| = O_p(1) \) (see Lütkepohl 1996, Result (4), sec. 9.11.2).

Remark 3.7. The factors \( D_t \) and \( F_t \) are allowed to be cross-correlated as well as serially correlated, although not perfectly collinear. For instance, the joint dynamics of \( Z_t \) could be described by a multivariate stationary ARMA. We are ruling out trending behaviours in \( D_t \) and \( F_t \), although our results can be suitably modified to accommodate trends.

Assumption 3.4 (regressors). For every \( i \), the \( T \times (K + S) \) matrix \( Z_i := [D_i, X_i] \) has full column rank. Moreover, for \( T, S \), satisfying \( 1/T + S/T = o(1) \), the matrix \( N^{-1} \sum_{i=1}^N \mathbb{M} z_i u_i' \mathbb{M} X_i \) has full rank with probability approaching one.

Remark 3.8. Assumption 3.4 requires enough cross-sectional heterogeneity of the \( X_i \) across units. Simple manipulations show that, with probability approaching one,

\[
\hat{S}_N = D_i' \left( \frac{1}{N} \sum_{i=1}^N \mathbb{M} z_i u_i' \mathbb{M} z_i \right) D_i = \frac{1}{N} \sum_{i=1}^N \mathbb{M} z_i u_i' \mathbb{M} z_i > 0,
\]

with \( \hat{S}_N \) as defined in (2.12).

Assumption 3.5 (loadings). \( \Gamma_1 \) and \( b_i \) are nonrandom such that \( \max_i \| \Gamma_1 \| < \infty \) and \( \max_i \| b_i \| < \infty \) and, as \( N \to \infty \),

\[
B_N := \frac{1}{N} \sum_{i=1}^N b_i b_i' \to B > 0.
\]
Additionally,
\[ A_N := \frac{1}{N} \sum_{i=1}^{N} \left( I_M - \Gamma_i \Psi_i^{-1} \Gamma_i' \frac{F.M.DF}{T} \right) \]
\[ b_i b_i' \left( I_M - \frac{F.M.DF}{T} \Gamma_i \Psi_i^{-1} \Gamma_i' \right) > 0 \] (3.6)
with
\[ \Psi_i := \Gamma_i' \frac{F.M.DF}{T} \Gamma_i + \Sigma_{W_i} \] (3.7)

**Remark 3.9.** Assumption 3.5 characterizes the behavior of the finite-dimensional matrices, namely \( A_N \) and \( B_N \), that play a key role in the behavior of the weighting matrix of the FWLS estimator. We differ from the usual weighted OLS estimation setup, as the FWLS weighting matrix can never be a diagonal matrix, unless \( b_i = 0 \), that is an \( M \times 1 \) vector of zeros.

Condition (3.5) implies that the factor structure (2.2) is strong, as defined in Pesaran and Tosetti (2011). This is commonly assumed in the factor models literature. However, there is a major difference between the factor models literature and this article. In that literature, the goal is to estimate the factors, and their loadings, and to do so one needs to eliminate the effect of the idiosyncratic component, by means of the averaging induced by the (static or dynamic) principal components. In this article, the aim is the opposite, namely to get rid of the cross-sectional heterogeneity in the sample. Finally, our results are similar to Pesaran (2006) and Bai (2009). The implications contain a weakly exogenous component, and in this respect, we are similar to Pesaran (2006) and Bai (2009). The implications from generalizing this assumption, in particular when considering dynamic panels in which one element of \( X_t \) represents the lagged dependent variable, are briefly discussed in Section G of the supplementary materials.

**Remark 3.10.** The technical condition (3.6) is used in the proof of Theorem 4.2. As shown in Appendix O.2, Lemma O.2(iv), supplementary materials the matrices in brackets are of full rank. Hence, (3.6) will be satisfied when there is enough cross-sectional heterogeneity in the sample. Finally, our results will not change if loadings are assumed random and cross-sectionally independent from other parameters.

**Assumption 3.6 (independence).** The \( (d_{ip}, f_{ij}), v_{ik}, e_{ju} \) are mutually independent for every \( i, j \) and \( t, s, u, \) and \( p, q, k \).

**Remark 3.11.** We are not allowing for any correlation between any entries of \( e_j \) and \( X_i \). This rules out the possibility that \( X_i \) contains a weakly exogenous component, and in this respect, we are similar to Pesaran (2006) and Bai (2009). The implications from generalizing this assumption, in particular when considering dynamic panels in which one element of \( X_t \) represents the lagged dependent variable, are briefly discussed in Section G of the supplementary materials.

**Remark 3.12.** Assumptions 3.2, 3.3, and 3.6 and Remark 3.6 imply that \( T^{-1} X_i'X_i \xrightarrow{p} \Sigma_{X_t} > 0 \) and \( T^{-1} X_i' \omega_d X_i \xrightarrow{p} \Sigma_{X_t, \omega_d X_t} > 0 \), for every \( i \). Hence, \( \left\| (T^{-1} X_i'X_i)^{-1} \right\| = O_p(1) \) and \( \left\| (T^{-1} X_i' \omega_d X_i)^{-1} \right\| = O_p(1) \).

**Remark 3.13.** Assumption 3.6 implies that sometimes \( \Psi_i \) and \( e_i \) can be interchanged with \( V_i \) and \( e_i \) without affecting the results; that is, they give rise to the same limits. For instance, \( \Sigma_{V_i} V_i = \Sigma_{V_i'} V_i, \Sigma_{V_i} e_i = \Sigma_{e_i'} e_i, \) and \( \Sigma_{V_i} e_i = \Sigma_{V_i'} e_i \).

### 4. Methodology: Asymptotics

The following two theorems enunciate the asymptotic properties of the OLS, WLS, and FWLS estimators, respectively, where \( d \rightarrow \) denotes convergence in distribution.

**Theorem 4.1.** When Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6 hold, as \( T \rightarrow \infty \), for any \( N \),

(i) (OLS estimator)
\[ T^{1/2} \left( \Sigma_i^{\text{OLS}} \right)^{-1/2} \left( \hat{\beta}_i - \beta_i - \tau_i^{\text{OLS}} \right) \xrightarrow{d} N(0, I_K) \]

where the bias term satisfies
\[ \tau_i^{\text{OLS}} := (\chi_i' \chi_i)^{-1} \chi_i' F.b_i \rightarrow p \tau_i^{\text{OLS}} := \Sigma_{\chi_i' \chi_i}^{-1} \Sigma_{\chi_i' \epsilon} b_i \] (4.1)

(ii) (WLS estimator)
\[ T^{1/2} \left( \Sigma_i^{\text{WLS}} \right)^{-1/2} \left( \hat{\beta}_i^{\text{WLS}} - \beta_i \right) \xrightarrow{d} N(0, I_K) \] (4.2)

where the asymptotic covariance matrix equals
\[ \Sigma_i^{\text{WLS}} := \Sigma_{\chi_i' \chi_i}^{-1} \Sigma_{\chi_i' \epsilon} \] (4.3)

**Remark 4.1.** The OLS (and the WLS) estimator is functionally independent of \( N \), which plays no role in the asymptotic analysis of Theorem 4.1. The OLS estimator is affected by a first-order bias. To simplify the exposition, we centered the OLS estimator around the sample bias \( \tau_i^{\text{OLS}} \) as opposed to the population bias \( \tau_i^{\text{OLS}} \). This has no relevant consequence as the scope of part (a) of the theorem is simply to bring attention to the existence of the bias plaguing the OLS estimator. Thus, the OLS estimator is asymptotically unbiased if either \( b_i = 0 \) or \( \Gamma_i = 0 \) or, alternatively, for diagonal \( \Sigma_{\chi_i' \epsilon} \) as well as with \( \Gamma_i \) and \( b_i \) satisfying \( \gamma_{ip} b_{ip} = 0 \) for every \( l \) and \( i \).

**Remark 4.2.** To gauge the intuition behind our WLS estimator, suppose that \( S = 0 \) (no common regressors), and \( M = K = 1 \) (one latent factor and one heterogeneous regressor) where the common factor vector equals the unit vector, that is, \( \mathbf{F} = \tau_{i} \), implying that model (2.1) can be written as
\[ \gamma_i = X_i \beta_i + u_i \text{ with } u_i = b_i \tau_{i} + \epsilon_i \text{ and } X_i = \gamma_i \tau_{i} + v_i \] (4.4)

Then, \( S_i = B_i \tau_{i} + I \) and \( S_i = \gamma_i \tau_{i} + v_i \) where \( N^{-1} \sum_{i=1}^{N} b_i^2 \rightarrow B = 1 \) and we set \( E\epsilon, \epsilon' = \Sigma_i = I_K \). By the Sherman–Morrison theorem (see Lemma L.2 in the supplementary materials) and some algebraic steps,
\[ S_i^{-1} \tau_{i} = \frac{\tau_{i}}{1 + \tau_{i}^2} = \frac{\tau_{i}}{1 + \tau_{i}^2}, \] (4.5)
that is, each element of the vector $S_1^{-1}X_i$ converges to zero at the fast rate $O(T^{-1/2})$. This fast rate is the key ingredient driving the asymptotic properties of the WLS and FWLS estimators. In particular, examining the random component (which involves $u_i$) of our WLS estimator $\hat{\beta}_{\text{WLS}}^i = \beta_{\text{WLS}}^i + (X_iS_1^{-1}X_i)^{-1}X_iS_1^{-1}u_i$, one obtains, by (4.5),

$$X_iS_1^{-1}u_i = (\gamma_iT + v_i)^T S_1^{-1} (b_iT + \epsilon_i) = v_i^T S_1^{-1} \epsilon_i + (b_iT + \gamma_iT)^T + (b_iT + \gamma_iT)^T (1 + T) \epsilon_i \epsilon_i^T + O_p(1 + T)$$

$$= v_i^T S_1^{-1} \epsilon_i + o_P(1 + T)$$

under our assumptions.\(^{10}\) Therefore, $X_iS_1^{-1}u_i$ is asymptotically equivalent to $v^T \epsilon_i$ implying that the weighting matrix $S_1^{-1}$ annihilates (asymptotically) the effect of the latent factors $F = \epsilon_i$. This result demonstrates how the WLS estimator $\hat{\beta}_{\text{WLS}}^i$ is asymptotically unbiased, because $v^T \epsilon_i = O_p(T^{-1/2})$ under our assumptions.

Considering now the general case $\Xi_i \neq 0$, by the same steps outlined above one obtains $X_iS_1^{-1}u_i = v_i^T \Xi_i^T \epsilon_i + O_p(1)$ and $X_iS_1^{-1}X_i = v_i^T \Xi_i^T v_i + O_p(1)$, implying that the WLS estimator is also asymptotically equivalent to the (unfeasible) GLS estimator, derived under the assumption that $\Xi_i$ is known.

The weighting matrix $S_i$ represents one of the many suitable choices for a weighting matrix. The limiting properties of the WLS estimator remain unchanged if one replaces $B$ with any other nonsingular positive definite matrix with bounded eigenvalues, although $B$ emerges when considering the FWLS. In our example above, we can replace $B = 1$ with any other positive constant, and obtaining identical limiting properties of the WLS estimator. In contrast, the matrix $\Xi_i$ is critical in order to achieve efficiency, as we have just outlined.

Further details are relegated to Section D of the supplemental materials, where we extend the above arguments to the case of generic latent factors $F$ when observed common regressors are allowed for ($S > 0$), and when $M$ and $K$ can be greater than one.

**Remark 4.3.** The assumption that the same latent factors $F$ enter into $X_i$ and $u_i$ is without loss of generality. In fact, assume $u_i = Gb_i + \epsilon_i$, with the rows of $G$ correlated but not identical to the rows of $F$. Then the bias takes the form

$$t_i^{\text{OLS}} = \Sigma_{X_iX_i^{-1}} - \Sigma_{F \Phi_T} \Phi_T G b_i,$$

exploiting the decomposition $G = \Phi_T G + \Phi_T G$, where $\Phi_T := I_T - \Phi_T$. Hence, the bias will only be nonzero due to the portion of $G$ correlated with $F$. The same consideration applies to the FWLS estimator. In Appendix F of the supplementary materials, we explore more in detail the implications of having different, yet correlated, factor structures for regressors and innovations.

We now present the main result of the article.

**Theorem 4.2.** When Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6 hold:

(i) As $1/T + T/N \rightarrow 0$,

$$\hat{\beta}_{\text{WLS}}^i \rightarrow^p \beta_i,$$

(ii) As $1/T + T^2/N \rightarrow 0$, then

$$\left( \left( \left( \left( V_i \Sigma_{X_iX_i^{-1}} V_i \right)^{-1/2} \left( V_i \Sigma_{X_iX_i^{-1}} V_i \right)^{-1/2} \left( V_i \Sigma_{X_iX_i^{-1}} V_i \right)^{-1/2} \left( V_i \Sigma_{X_iX_i^{-1}} V_i \right)^{-1/2} \right) \right)^{-1} \right) \rightarrow^d N(0, I_K),$$

where

$$C_N := D_N^2 C_N D_N, \quad C_N := \frac{1}{N} \sum_{i=1}^N (\hat{\xi}_i + \Theta_i)$$

with

$$\Theta_i := \mathbb{E} \left[ V_i \Sigma_{X_iX_i^{-1}} \gamma_i \Sigma_{F \Phi_T} b_i b_i^T \Sigma_{F \Phi_T} \Gamma_i \Sigma_{X_iX_i^{-1}} V_i \right].$$

(iii) For any integer $\tau < T$ satisfying $1/\tau + \tau / T^2 + T^2 / N \rightarrow 0$

$$\left( \left( \left( \left( V_i \Sigma_{X_iX_i^{-1}} V_i / T \right)^{-1} \right)^{-1} \right)^{-1} \right)^{-1} \rightarrow^p 0_{K \times K}$$

and

$$\left( \left( \left( \left( \hat{\xi}_{\phi \phi} + \sum_{h=1}^\tau \hat{\ell}(h, \tau) \left( \hat{\xi}_{\phi \phi} \right) \right)^{-1} \right)^{-1} \right)^{-1} \rightarrow^p 0_{K \times K},$$

setting $\hat{\ell}(j, \tau) := 1 - |h/(\tau + 1)|$, and $\hat{\xi}_{\phi \phi} = T^{-1} \sum_{t=h+1}^{\tau+m} \hat{a}_{it} \hat{a}_{jt} \hat{a}_{it} \hat{a}_{jt} \omega_{it} \delta_{jt}^{-1}$, where $\hat{a}_{it}$ denotes the $t$th element of $(T \times 1)$ vector $\hat{a}_i := \tilde{S}_N^{-1/2} a_i$, with $\tilde{a}_{i}^{\text{FWLS}} := y_i - \chi_i \hat{\beta}_i$ and $\omega_{it}^{-1}$ denotes the $t$th row of the $(T \times K)$ matrix $\hat{W}_i := \tilde{S}_N^{-1/2} \chi_i$.

**Remark 4.4.** To simplify the exposition and the proofs, we have not differentiated the conditions required for consistency and asymptotic normality of the FWLS estimator, such as the moment conditions on the regressors and the error term. However, different rates for $N$ and $T$ are required for consistency and asymptotic normality of the FWLS estimator.

**Remark 4.5.** A mechanism, similar to the one described in Remark 4.2 for the WLS estimator, applies to the FWLS in terms of bias-reduction, in turn granted by annihilating the $\Phi_T$, as long as $S_i^{-1}$ is replaced by any matrix of the form $(\Phi_T \Phi_T + C)^{-1}$, where $A$ and $C$ are, as minimum conditions, nonsingular matrices (in general, functions of $N$ and $T$) with uniformly bounded eigenvalues. The most challenging aspect of the proof to Theorem 4.2 is to establish that $\tilde{S}_N^{-1}$ represents (asymptotically) one of those matrices, where in particular its $C$ matrix is approximately (in a precise sense defined below) behaving like $N^{-1} \sum_{i=1}^N \hat{\xi}_i$, as a byproduct of the averaging implicit in the construction of $\tilde{S}_N$ (by (2.12)). Thus, such $C$ matrix can never be equal to $\Xi_i$, except (generically speaking) when $\Xi_i = \Xi$, an homogeneity case across units of the $\xi_i$. Therefore, the FWLS estimator is not efficient in general, as evinced by the form of the asymptotic covariance matrix in Theorem 4.2.
Remark 4.6. A multistep procedure can be envisaged that improves the precision of the FWLS and, under homogeneity of the covariance matrix of the error, achieves an estimator with an asymptotic distribution arbitrarily close to the asymptotic distribution of the WLS estimator. We shall call this procedure the iterated-FWLS estimator. The first step starts from the FWLS estimator, which we now denote as \( \hat{\beta}_{1}^{(1)} \) for simplicity. We then construct the associated residuals \( \hat{u}_{i}^{(1)} := y_{i} - x_{i}^{'}\hat{\beta}_{1}^{(1)} \). Recall that \( \hat{\beta}_{1}^{(1)} = \hat{\beta}_{1}^{\text{FWLS}} \) is a consistent estimator for \( \beta_{1} \), and thus, the residuals \( \hat{u}_{i}^{(1)} \) are close to the error \( u_{i} \), asymptotically. The second step entails constructing \( \hat{\delta}_{N}^{(1)} := N^{-1} \sum_{i=1}^{N} \hat{u}_{i}^{(1)} \hat{u}_{i}^{(1)'} \) and using it to obtain \( \hat{\beta}_{1}^{(2)} := \left( \mathbf{X}_{i}^{'} \left( \hat{\delta}_{N}^{(1)} \right)^{-1} \mathbf{X}_{i} \right)^{-1} \mathbf{X}_{i}^{'} \left( \hat{\delta}_{N}^{(1)} \right)^{-1} y_{i} \).

In general the \( h \)th step entails constructing \( \hat{\beta}_{i}^{(h)} := \left( \mathbf{X}_{i}^{'} \left( \hat{\delta}_{N}^{(h-1)} \right)^{-1} \mathbf{X}_{i} \right)^{-1} \mathbf{X}_{i}^{'} \left( \hat{\delta}_{N}^{(h-1)} \right)^{-1} y_{i} \), where \( \hat{\delta}_{N}^{(h-1)} \) is obtained based on \( \hat{u}_{i}^{(h-1)} := y_{i} - x_{i}^{'}\hat{\beta}_{i}^{(h-1)} \).

5. Numerical Results

5.1. Monte Carlo Analysis

We conduct a set of Monte Carlo experiments to evaluate the performance of our asymptotic results for the (iterated) FWLS estimator in finite samples, and compare it with the WLS and OLS estimators, and CCE estimator of Pesaran (2006), the latter being the most commonly used estimator for model (2.1).

We construct various data-generating processes (DGP) that provide an exhaustive range of situations of interest.

**Design.** We generate the data according to

\[
y_{it} = \alpha_{0t} + \beta^{0} a_{ixi} + b_{ix} f_{i} + \varepsilon_{it},
\]

where \( K = 1 \), and the single heterogeneous regressor satisfies \( x_{it} = 0.5 + y_{i0} f_{i} + v_{it} \), implying the single observed common factor \( f_{i} \) equal to 1 for all observations. The parameters are generated according to \( \beta_{0} \sim NID(1, 1.04) \), and thus, heterogeneous across units, where \( NID \) means iid normally distributed, \( \alpha_{0} = -0.5 \) for \( i = 1, \ldots, N/2 \) and \( \alpha_{0} = 0.5 \) for \( i = N/2 + 1, \ldots, N \), the M latent common factors according to \( f_{i} = 0.5 f_{i-1} + 0.5 \eta_{ij}, \) with \( (\eta_{ij} \sim NID(0, 1/2)) \), and the error terms to the \( y_{it} \) and \( x_{it} \) according to \( \varepsilon_{it} = \rho_{\varepsilon} \varepsilon_{it-1} + \eta_{ij}, \) with \( \eta_{it} \sim NID(0, 1/2)) \), and to \( v_{it} = \rho_{v} v_{it-1} + \eta_{it}, \) with \( \eta_{it} \sim NID(0, 1/2)) \), with \( \rho_{\varepsilon} \sim UID(0.05, 0.95), \rho_{v} \sim UID(0.05, 0.95), \) and \( \sigma_{0}^{2} \sim UID(0.5, 1.5), \) where \( UID \) means iid uniformly distributed.

Finally, regarding the loadings, we consider four cases:

- **DGP1.** \( M = 2, b_{i0} \sim NID(\frac{1}{2}, 0.2I_{2}), \)
- \( y_{i0} \sim NID(0.5, 0.5I_{2}), \) and mutually independent.

- **DGP2.** \( M = 2, b_{i0} \sim NID(\frac{1}{2}, 0.2I_{2}), \) \( y_{i0} = b_{i0}, \)

- **DGP3.** \( M = 4, b_{i0} \sim NID(\frac{1}{2}, 0.2I_{2}), \)
- \( y_{i0} \sim NID(0.5, 0.5I_{4}), \) and mutually independent.

- **DGP4.** \( M = 2, b_{i0} \sim NID(0, 0.2I_{2}), \)
- \( y_{i0} \sim NID(0.5, 0.5I_{2}), \) and mutually independent.

The meaning of the above DGPs is as follows: DGP1 is the benchmark case; DGP2 represents the case of (perfect) cross-correlation between the loadings in the regressor and in the error term; DGP3 represents the case when \( M = 4 > K = 1 \), that is when the number of latent factors exceeds (substantially) the number of regressors; DGP4 represents the case when the factors affecting the error term are not pervasive. The DGPs here follow the Monte Carlo design of Pesaran (2006). We have considered alternative Monte Carlo designs, such as the one in Karabiyik, Urbain, and Westerlund (2019), with extremely similar results. (Details are available upon request.)

**Results.** We consider 1000 Monte Carlo replications with sample sizes \((N, T) \in \{60, 200, 600\} \times \{30, 100, 300\} \), where \( N > T \). Under DGP1 of Table 1, all estimators work, in terms of absolute bias (expressed as the percentage change, in absolute value, of the estimate from the true value) and root-mean-squared-error, except for OLS. As predicted by our theory, the FLWS works particularly well when \( N \) large, especially when \( N \) and \( T \) are of comparable magnitude, and is commensurate to the CCE. The performance of the WLS and GLS remains similar to one another throughout all the DGPs. Tables 2, 3, and 4 show the performance of the estimators in the realistic scenarios of interest indicated by DGP2, DGP3, and DGP4. When the loadings to the latent factors, \( b_{i0} \) and \( y_{i0} \), are (perfectly) cross-correlated (Table 2), the FLWS continue to perform well, whereas the CCE
becomes severely biased. The same pattern is observed (Table 3) when the number of latent factors $M$ is much larger than the number of observed regressors ($K$), and even when the factor structure in the error term of (5.1) is weak (Table 4). The last three columns of each table report the empirical sizes of the $t$-ratios, defined as $\sqrt{T}(\hat{\beta}_m - \beta_0)/\text{SE}_{\hat{\beta}_m}$ for the FWLS estimator ($\text{SE}_{\hat{\beta}_m}$ denotes the standard error of $\hat{\beta}_m$ as from Theorem 4.2-(iii)), which appear close to the nominal sizes of 1%, 5%, and 10%, across all tables in most cases, with some deterioration occurring when $T$ remains small.

### 5.2. Empirical Application: Capital Accumulation and GDP Growth

We investigate the empirical relationship between investment in physical capital and long-run economic growth across countries, a topic of paramount importance for quantifying the effectiveness of development policy, and also representing a way to evaluate economic growth theories. Endogenous growth models, the so-called AK models (see, Romer 1986, among others), predict that an increase in investment results in an increase in physical capital and long-run economic growth across countries.
a permanently higher growth rate, implying a reduced form equation for per capita GDP growth equal, in its simplest form, to

$$\Delta y_{it} = \theta_1 + \theta_2 x_{it} + \beta_1 \Delta x_{it} + u_{it},$$  \hspace{1cm} (5.2)

indicating by $y_{it}$ the logarithm of investment as a share of gross domestic product (GDP) for country $i$ in year $t$, and by $\Delta y_{it}$ the logarithm of output per worker for country $i$ in year $t$, with $\Delta y_{it}$ denoting the annual per capita GDP growth. Model (5.2) implies a steady-state GDP growth equal to $\Delta y_1 = \theta_1 + \theta_2 x_i$. In contrast, exogenous growth theories, variation of the celebrated Solow model (see Mankiw, Romer, and Weil 1992, among others), postulates that the long-run growth rate of output per worker is only determined by the exogenous technical progress but predict a positive correlation between investment and the level of per capita GDP, ruling out any association between investment and steady-state growth rates. These theories can be expressed by the reduced form equation

$$y_{it} = \theta_0 + \theta_1 t + \beta_1 x_{it} + u_{it},$$  \hspace{1cm} (5.3)

implying the constant steady-state GDP growth $\Delta y_1 = \theta_1$. Now, a permanent increase in the investment-GDP ratio $x_{it}$ predicts a higher level of output per worker $y_{it}$ along the steady state growth path, but affects growth only during the transition to the new steady state.

The existing empirical evidence is mainly focused on the OECD countries and is mixed at best, with many studies concluding that there is a weak association between investment rates and long-run growth rates, casting doubt on the effectiveness of physical capital accumulation as a source of long-run economic growth (see Jones 1995; Bond, Leblebicioglu, and Schiantarelli 2010, among others). Most of the empirical studies consider the case of constant slopes in the estimated regression, implying that one can only learn about average effects, leaving out the substantial heterogeneity existing among countries. This is a consequence of these studies using cross-sectional and panel regressions, that unavoidably do not allow identification of the heterogenous trend slopes. Moreover, capital accumulation and GDP growth are no doubt jointly determined, implying that standard techniques such as OLS are affected by simultaneous equation bias. Our estimation procedure, based on the FWLS approach, tackles both the problem, offering a substantial improvement over the existing econometric techniques used so far in the empirical growth literature.

We adapt Bond, Leblebicioglu, and Schiantarelli (2010) (see their eq. (7)) and postulate the following linear relationship between the logarithm of GDP per capita $y_{it}$, and the logarithm of the investment share to GDP $x_{it}$, which nests both the endogenous and the exogenous growth model specifications (5.2) and (5.3):

$$\Delta y_{it} = \theta_1 + \theta_2 x_{it} + \beta_1 \Delta x_{it} + \beta_2 \Delta x_{it-1} + u_{it},$$  \hspace{1cm} (5.4)

or, equivalently, $\Delta y_{it} = \alpha_i d_t + x_{it} \beta_t + u_{it}$ setting $d_t = 1, x_{it} = (x_{it}, \Delta x_{it}, \Delta x_{it-1})'$. Importantly, (5.4) allows for heterogeneity across countries of the regression coefficients, a crucial feature given that we are using data of both OECD and Non-OECD countries, for which the assumption of constant growth rate across countries is untenable. Model (5.4) implies that the GDP per worker, for country $i$, grows at the country-specific rate of $\Delta y_{it} = \theta_1 + \theta_2 x_{it}$ in steady state, nesting both the endogenous (when $\theta_2 = 0$) and the exogenous (when $\theta_2 \neq 0$) cases. Therefore, the long-run effect of an increase in the investment share on the growth rate of output per worker, defined as the growth effect, is $\theta \Delta y_{it} / \Delta x_{it} = \theta_2$, and the the long-run effect of an increase in the investment (log) share on the (log) level of output per worker, defined as the level effect, is given by $\theta y_{it} / \Delta x_{it} = (\beta_1 + \beta_2)$.

We use the data extracted from both the Penn World Table 6.3 (PWT 6.3) regarding real GDP per worker and the share of total gross investment in GDP, both measured in constant-price international dollars, for $N = 151$ countries (OECD and Non-OECD countries), yearly observations from 1970 to 2007. Estimation of model (5.4) is challenging due to the strong degree of persistence, both time series and cross-sectional. Taking first-difference of the (log) GDP per capita, as in (5.4), mitigates the first concern, whereas this is less concerning for the (log) investment share $x_{it}$ as it represents a fraction. Figure 1, left panel, reports the $p$-values of the Augmented Dickey–Fuller test, across countries, for the log GDP (black line) and its first-difference (red line), together the 5% level (blue line), showing the effectiveness of the taking the first-difference. We did not take the first-difference of the investment share $x_{it}$ as the assumption of stationarity is plausible given that $x_{it}$ represents a ratio bounded between zero and one.

Regarding the degree of cross-sectional dependence, Figure 1, right panel, reports the largest 32 eigenvalues in descending order (normalized by setting the largest equal to 1) of $y_{it}$ (black line) and $x_{it}$ (red line), with both exhibiting a prominent factors structure. Moreover, the two sets of estimated factors, corresponding to the eigenvalues of Figure 1 (right panel), appear strongly cross-correlated, as indicated by their canonical correlations, the first five of which are equal to 0.976, 0.905, 0.875, 0.799, and 0.701 in descending order, denoting substantial cross-sectional dependence.

Therefore, it is very likely that estimation by OLS of (5.4) is affected by (first-order) bias, making this an ideal setup for our FWLS estimator, as it allows the presence of a factor structure (with an arbitrary number of latent factors), both in the dependent and independent variates, together with arbitrary heterogeneity of the regression coefficients.

16 Bond, Leblebicioglu, and Schiantarelli (2010) augment (5.4) with lagged $\Delta x_{it}$ to enrich the short term dynamics of the model, but the magnitude of the corresponding estimated parameters turns out to be negligible in terms of growth and level effects, and thus, we consider the more parsimonious model (5.4). The presence of the lagged $\Delta x_{it}$ in (5.4) accounts for the dominant part of this dynamic effect.

17 We excluded all countries with gaps over the chosen data period, countries for which oil production is dominant, and countries with reported negative gross investment over some years.

18 Use of the Augmented Dickey–Fuller test shows that for the investment share the $p$-value is below 0.2 for about half of the countries whereas for log GDP frequency of countries with a $p$-value below 0.2 is only about 15%.

19 Canonical correlations, introduced by Hotelling (1936), permit to quantify the degree of dependence between two sets of variates, and are defined as the minimum eigenvalues of the matrix $(A'A)^{-\frac{1}{2}}(A'B)(B'B)^{-\frac{1}{2}}B'A(A'A)^{-\frac{1}{2}}$, denoting by $A, B$ the samples of data for the two sets of variates, respectively, of dimension $T \times K_A$ and $T \times K_B$, assuming zero mean for simplicity.
Table 5 and Figure 2 report our empirical results. Generally speaking, we find robust evidence of a positive, often significant, relationship between investment and GDP growth, on average across countries. Figure 2 shows that there is a substantial degree of heterogeneity, and asymmetry, across countries, with the estimated cross-sectional densities appearing all shifted toward positive values (positive skewness) of the growth and level effects, except for the level effect of the OECD countries which appears to have negative skewness. The FWLS estimates reported in the top panel of Table 5 show clearly that some differences arise from comparing OECD and Non-OECD countries, with the growth effect being stronger with the former group.
A factor structure better captures the dependence, especially across the asymptotically normal when both the cross-section and the regression error. The FWLS estimator is consistent and panels with a common factor structure in both the regressors and the error are correlated; fifth, it can deliver efficient estimation under suitable conditions on the covariance matrix of the idiosyncratic error. Our results are corroborated by a set of Monte Carlo experiments and illustrated by an empirical application that investigates the effect of investment on GDP growth. Several generalizations, such as dynamic panels, cross-section regressions with time-varying coefficients, and different factor structures for regressors and error, are described in the supplementary materials.

### 6. Concluding Remarks

This article proposes weighted least square estimation for linear panels with a common factor structure in both the regressors and the regression error. The FWLS estimator is consistent and asymptotically normal when both the cross-section N and time series T dimensions diverge to infinity where under the same circumstances, the OLS is first-order biased. In summary, the FWLS estimator exhibits five main properties: first, it permits carrying out standard inference on the regression coefficients because based on conventional limiting distributions; second, it does not require any knowledge of the exact number of latent factors, or even an upper bound of such number; third, it is computationally easy to handle without invoking any nonlinear numerical optimizations; fourth, it works regardless of whether the regressors and the error are correlated; fifth, it can deliver efficient estimation under suitable conditions on the covariance matrix of the idiosyncratic error. Our results are corroborated by a set of Monte Carlo experiments and illustrated by an empirical application that investigates the effect of investment on GDP growth. Several generalizations, such as dynamic panels, cross-section regressions with time-varying coefficients, and different factor structures for regressors and error, are described in the supplementary materials.

### Supplementary Materials

Supplementary material contains additional materials, such as notation, extended literature review, the intuition behind the estimator and generalizations of the method, and the proofs of main theorems with the required lemmas.

### Acknowledgments

The authors are extremely grateful to Joakim Westerlund, to the Associate Editor, and to the anonymous referees for comments that greatly improved the article.

### References


